Review of: “Scientific and Technical Basis of the Numeric Nutrient Criteria for Montana’s Wadeable Streams and Rivers: Addendum 1”

Peer Review Questions

1. *Approach* - Montana’s approach combines stress-response studies and ecoregional reference distributions to derive numeric nutrient criteria for Montana’s wadeable streams (not really two approaches as identified in review question #1). The approach relies principally on stress-response studies both within the ecoregion and in nearby regions if the reference distributions of TN and TP of the 2 regions are similar. If no stress-response studies were deemed relevant, then MDEQ relied on the reference distribution (e.g., 75^{th} %ile of reference for TP in Absaroka-Gallatin ecoregion 17i). As a further condition on the approach, MDEQ also keeps the N:P ratio in the criteria in a “Redfield range” so that the N:P ratio is unlikely to deviate far from the Redfield ratio, or far from the ecoregional reference if the reference deviates far from Redfield (e.g., Absaroka-Gallatin). The stress-response studies include both experimental nutrient enrichment studies and empirical modeling studies based on monitoring data. Overall, I find the whole approach compelling because it makes effective use of available and relevant information.

   I think the approach would be strengthened by increased use of Montana’s own monitoring data to develop ecoregion-specific empirical models of benthic chl-a response to nutrient enrichment. Such stress-response studies were indeed used if they were available as separate reports or publications, but there is no systematic application of Montana’s data to derive empirical models or confirm the proposed criteria. These models could be used to confirm, refute, or adjust the criteria developed, and would further strengthen the criteria.

   The second part of EPA’s review question, starting with “Please provide documentation on any identified ranges…” is out of line. It is not a reviewer’s task to develop a compendium of alternative methodologies, advantages, disadvantages, data, tools, etc.

2. *River Breaks* – The description of nutrient conditions in the river breaks is plausible, and the river breaks region seems similar (though maybe less extreme) than other badlands ecoregions in the Northwestern Great Plains (including badlands regions of the Dakotas). However, these regions are not familiar to many persons steeped in the Eastern Forest Biome stream paradigms. Accordingly, MDEQ should provide more documenation for the assertions made about the River Breaks. Also, would the same considerations apply to the Little Missouri badlands and the Missouri River Breaks? Evidence could include: published stream studies of badlands-type regions, provided they are similar to the River Breaks; N and P content of soils and geological formations in the River Breaks and similar regions (for example, I have found from EPA’s ecoregion descriptions that the Cretaceous Hell Creek Formation occurs in several of the badlands/breaks ecoregions). Land use/land cover and population density could help show that the breaks and badlands are no different in land use than other Northwestern Great Plains

regions, and perhaps even lower population density and less alteration of land use/land cover than other parts of the Northwestern Great Plains. Early historic descriptions of the regions and their streams are also highly useful, if available.

What is missing from the River Breaks criteria is protection of downstream waters, the large rivers and reservoirs. For example, parts of the region drain into Fort Peck Reservoir. Nutrients in the Breaks streams could contribute to eutrophication of the reservoir. If there are no criteria for the Breaks region, then we could envision the following scenario: In the absence of criteria, the area could become a magnet for large industrial feedlots because no nutrient removal would be required. What happens when hundreds of feedlots drain into the Breaks and on into Ft Peck reservoir? Although criteria may not be required to protect the aquatic life in the streams, they may be required to protect downstream waters.

3. *20% exceedance* – As far as I understand, the 20% exceedance rule means that no more than 20% of single measurements may exceed the nutrient criteria concentrations, or that the site may exceed a criterion up to 20% of the time. Since Montana's proposed criterion is based on single measurements, it is reasonable to expect that some short-term variation above the nutrient criterion concentration will not result in excess chlorophyll. The other alternative is to frame the criteria in terms of an annual average (say, geometric mean) as EPA did for the Florida nutrient criteria. A central tendency measure, geometric or otherwise, also allows for some short-term high concentrations as long as the central tendency is not exceeded. Montana's is basically the same, but based on a percentile of individual measurements.

   The 20% frequency was based on analysis of the Clark Fork River, which shows that for the Clark Fork, the 20% criterion would work well. The problem is that the Clark Fork is a single basin in a restricted set of subecoregions, so we don't have empirical evidence whether 20% would apply to the rest of the state as well. Using Montana's existing monitoring data, I think it may be possible to repeat some of the Clark Fork analysis on other streams throughout the state to confirm or refute the 20% estimate. The Clark Fork data presents another opportunity as well: testing the entire nutrient assessment approach to determine if the actual error rates match with the desired alpha and beta of 0.25 and 0.30. <u>Recommendation</u>: the 20% exceedance rule seems reasonable and has empirical evidence to support it, but would be strengthened by additional analysis from other regions of the state.

4. *125 mg/m^2 chl a* –Benthic chl-a is clearly the most consistent response indicator to nutrients in wadeable streams, as shown by many studies, cited in the MT documents and elsewhere. Benthic macroinvertebrates, while associated with both nutrients and chl-a, have so far proved unsuccessful as a reliable response indicator to nutrient enrichment, as demonstrated by EPA's attempt to develop nutrient criteria for Florida streams. Montana has a rich tradition in monitoring benthic chl-a as well as benthic diatoms, and is making effective use of that tradition for developing nutrient criteria.
   Derivation of the threshold – The threshold was derived from literature values, observations of

streams in the MT ecoregions, an acceptability survey, and a nutrient enrichment study. For example, Welch et al. (1989; cited in MT docs) considered "Nuisance biomass levels" to be in the range 100 – 150 mg/m^2 chl-a. Other values are similar (Biggs 2000: mesotrophy is in the range 60-200 mg/m2; Dodds et al. 2002 [CJFAS 59:865-874]: 125 mg/m^2 is "high end" of chl-*a*). Surveys are context-specific, in that people will identify unacceptable conditions as those that they are not accustomed to seeing. Unaccepatibility thresholds are subject to shifting baselines: if the persons surveyed are accustomed to seeing eutrophic conditions, only hypereutrophy would be identified as unacceptable. Finally, the dose-response study suggested that synoptic reach-average benthic algae in the range 87 – 127 mg/m2 chl a resulted in unacceptable DO at the end of the growing season. These results would suggest that 125 mg/m2 is at or uncomfortably close to a value that could cause fish community degradation due to DO, and for mountain and transitional streams, the chl-a threshold should be lower.
Statewide use - First, it is unclear whether MT plans to use 125 or 150 mg/m^2 as the chl-a standard. Some regions have 125, others 150. As with the nutrient criteria themselves, it may be more appropriate to have chl-a criteria better adjusted to the ecoregions. For example, the expectation for mountain and foothill-transitional ecoregions is that streams are oligotrophic and coldwater, supporting Montana's famous trout fisheries. Given that 125 mg/m2 is in the range of "nuisance", well in "mesotrophy" and has been demonstrated to cause DO problems in Montana, this value is probably too high for mountain and foothill streams. I have no problem with higher values for Plains ecoregions.

5. *Reach-specific criteria* – Two methods for reach-specific criteria are proposed: empirical determination based on pre-defined natural conditions (in this case, dam operations), and ecoregional flow-weighted criteria for streams receiving input from more than a single ecoregion. Both of these approaches appear to be sound.

6. *Tests* – Montana's overall rationale for determining impairment, using both an exact binomial and the t-test, is well thought-out. However, the presentation was a bit confusing; I found I had to jump around between various parts of the 2011 document and its appendixes to understand the approach. The consideration of both significance and power, and the attempt to balance them, is especially encouraging, and shows MDEQ is concerned with both protection of the resource and prevention of unnecessary management. I do have some concerns:

    a. Are the effect size (0.15) and the critical exceedance rate (0.20) really double-counting the same thing? Effect size is a scientific determination that nutrient concentrations within 15% of each other (or within 5% of the criterion) are not meaningfully different in terms of response, so it protects against a statistically significant difference (which may be significant simply due to very large sample size) being declared an impairment when there is actually little chance of impairment for such a small difference. The exceedance rate essentially does the same thing: up to 20% of individual measurements can exceed the criteria, but chl-a will not exceed its criterion value. When both of these are used in the exact binomial, is it testing whether more than 20% of observations exceed the critical nutrient concentration plus 15%? If so, that would be double-counting. The Clark Fork data could be used to test/illustrate this issue empirically. Effect size is typically used in comparisons of central tendency, to protect against scientifically negligible differences being elevated to statistically significant differences simply due to large sample size. It is used most often in equivalence or noninferiority tests. I don't think effect size, as a % of the mean, is appropriate for the exact test, which does not use a mean.

b. Should the effect size be used in the t-test, especially for large or very large sample sizes?

General comments:

In view of several of the questions above, and different ways of calculating status of a streams reach, I was frequently confused whether the document was referring to instantanaeous measures, annual (growing season) maximum, or some measure of central tendency (mean, median, geometric mean, etc.) measure at one time (synoptic) at several sites on a reach, or a "sampling event average". I came to realize that Montana's proposed criteria only make sense in the context of individual measures, i.e., measurements of TN and TP are not to exceed the criterion more than 20%. Similarly, the chl-a criterion of 125 mg/m2 only makes sense as a maximum not-to-be-exceeded. However, it was not clear in the document how exceedance would be calculated. Critically, eventual measures of exceedance should match to the extent possible the way meaningful concentrations were calculated in the considerations to derive the criteria. Recommendation: spell out, with examples, of what is meant by single observations and different central tendencies mentioned in the documents, and which are used for the final criteria and for assessment.

Overall, I found Montana's approach sound and well thought-out. The devil, as always, is in the details: selection of chl-a values, derivation of critical exceedance rates, selection of effect size. Some of the quibbling on these values may never be resolved (including mine), and Montana needs to use best judgment supported by its analysis and other scientific results.